DOCUMENT RESUME

ED 243 947 TM 840 260

AUTHOR Mayberry, Paul W.
TITLE Analysis of Cross-Cultural Attitudinal Scale Translation Using Maximum Likelihood Factor Analysis.
PUB DATE Apr 84
NOTE 45p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Adults; *Attitude Measures; *Factor Analysis; Factor Structure; Japanese; Latent Trait Theory; *Maximum Likelihood Statistics; Multivariate Analysis; Statistical Significance; Test Bias; *Testing Problems; Test Theory; *Translation
IDENTIFIERS Inventory of Personal Investment (Braskamp Maehr)

ABSTRACT

Efforts to study the fidelity of translation of attitudinal scales into foreign languages have faltered due to the lack of powerful statistical tests to assess such transformations. This study uses a maximum likelihood factor analysis procedure to compare multivariate factor structures across subpopulations. The results showed that inconsistent measurement can be identified in the analysis of cross-cultural scale translations for attitudinal instruments. The maximum likelihood factor analysis procedure provided many advantages over the classical test theory and item response theory approaches for the detection of inconsistent items in attitudinal instruments. Despite its limitations, the technique was readily able to accommodate multidimensional data sets that were polychotomously scored. The procedure also provided a very rigorous statistical test that is rooted in factor analysis research and is a viable alternative to the more restrictive item response theory techniques. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED243947

ANALYSIS OF CROSS-CULTURAL ATTITUDINAL SCALE TRANSLATION USING MAXIMUM LIKELIHOOD FACTOR ANALYSIS

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Paul W. Mayberry

Measurement and Research Division
Office of Instructional Resources
307 Engineering Hall
1308 W. Green
University of Illinois at Urbana-Champaign
Urbana, IL 61801

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

P. W. Mayberry

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM840260

Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA, April 1984. Comments are welcome.

ANALYSIS OF CROSS-CULTURAL ATTITUDINAL SCALE TRANSLATION USING MAXIMUM LIKELIHOOD FACTOR ANALYSIS

Efforts to study the fidelity of translation of attitudinal scales into foreign languages have faltered due to the lack of powerful statistical tests to assess such transformation (Irvine & Carroll 1980, Katerberg, Hoy, & Smith, 1977). Traditional procedures for assessing the quality of translation are usually judgemental involving back translations into the original language or comparisons of responses by bilingual individuals. Both techniques lack sufficient rigor in addressing the dual question of adequate translation. Skilled translators can often produce reasonably acceptable back translations from inferior original versions by inferences and insightful guessing but this does not address the issue of cross-cultural "meaningfulness" of the instrument. The use of bilingual subjects ignores the fact that this sample is not representative of the monolingual individuals to which the researcher wishes to infer. On the other hand, the more measurement oriented procedures generally rely on classical test theory comparisons of item discriminations or difficulties in their attempt to identify insufficient translations. Yet the populations may vary on the latent traits being measured by the attitudinal instrument and which accordingly can not be identified by such procedures. Thus it is necessary to test both the quality of translation and the equivalency of latent structure for the groups of interest in order to evaluate cross-cultural translations of attitudinal assessment instruments.

Research in the achievement domain provides strong empirical support for the superiority of item response theory (IRT) over its classical test theory counterparts (Lord 1980, Wright 1977, Scheuneman 1979, Ironson & Subkoviak 1979). Hulin, Drasgow, and Komocar (1982) have generalized the item bias detection methods of IRT to assess scale translations. Their analysis centered

on the Job Descriptive Index (JDI) which is quite adaptable to IRT procedures. Yet the JDI is not typical of most attitudinal instruments in that it is essentially dichotomous in its item responses and has an extremely dominant first common factor (thus the instrument does not significantly violate the unidimensionality assumption required for IRT). While IRT does provide a rigorous analysis for the application of scale translation, its transfer to other attitudinal measures is questionable. Being such a powerful theory of measurement, IRT relies on a strong adherence to its underlying assumptions; in particular, unidimensionality. Unidimensionality of the latent space is not a realistic assumption with most attitudinal instruments which are usually designed to measure an entire range of constructs. Efforts to produce qualifying unidimensional scales often result in too few items per scale, which in turn distort item parameter estimates. Also, current parameter estimation procedures restrict the scoring of item responses to being dichotomous. Loss of information is inevitable in the transformation from a multiple response format to a binary classification of the item's content. Finally, sample size requirements for stable parameter estimates often restrict the application of IRT.

A technique that is more responsive to these attitudinal measurement considerations is a maximum likelihood factor analysis procedure comparing multivariate factor structures across various subpopulations, often referred to as SIFASP (Joreskog 1971). Drasgow and Kanfer (1983) were the first to propose using SIFASP to study "equivalence of measurement" with respect to attitude instruments. This approach readily accomodates polychotomous item responses while being able to converge for a smaller number of items and a smaller sample size. Also, the factorial model is able to function in a multidimensional latent space as well as perform simultaneous comparisons of two or more groups.

The SIFASP technique allows for factorial model comparisons in the testing

of various hypotheses concerning the interitem covariance matrices. These hypotheses postulate the factor structure for various subpopulations by making constraints on the parameter estimates and holding these estimates invariant across the subgroups. If the usual assumptions of the factor analysis model are made (i.e., unique variables are mutually uncorrelated and uncorrelated with the common factors), the variance-covariance matrix of the observed variables can be expressed as:

$$(1) \quad \hat{\Sigma} = \Lambda_y \Phi \Lambda_y' + D^2.$$

where Λ_y is the pattern matrix, Φ is the interfactor correlation matrix, and D^2 is a diagonal matrix of the uniquenesses.

The basic premise of this "equivalence of measurement" analysis is based on Meredith's (1964) application of Lawley's (1943-44) selection theorem with respect to the covariance matrix described in equation 1. Meredith demonstrated that if the factor analytic model holds in a parent population, then it also holds in any subpopulation formed by any selection variable, assuming that each unique variable is independent of that selection variable. This results in the factor pattern matrix (Λ_y) being <u>invariant across subpopulations</u> despite the selection variable (thus this variable does not necessarily have to be observed). The unique variances (D^2) are also invariant but the factor variance-covariance matrix (Φ) is affected by selection. As a result of Meredith's proof, the expected observed scores for individuals with identical factor scores (i.e., equal standing on the latent structure) can be translated as being equivalent regardless of group membership. Also, since the Λ_y and D^2 matrices are invariant across subpopulations, the differences between subpopulations affect <u>only</u> the factor

variance-covariance matrix. Therefore, the purpose of this study is two-fold with respect to applying the maximum likelihood factor procedure: (1) to assess the quality of scale translation for a multidimensional attitudinal instrument from an original to a target language, and (2) to determine the degree of equivalence of the latent structure for the cross-cultural subpopulations.

METHODOLOGY

Instrument

The theory of personal investment (Maehr 1984) postulates that individuals expend differing amounts of their time, talents, and resources in an attempt to maximize their overall satisfaction or feelings of importance. There are various personal incentives for investing oneself in some areas but not others. Six of these incentives were studied here as measured by the Inventory of Personal Investment, ${}_{I}P_{I}$ (Braskamp and Maehr 1983).

The rendition of an attitudinal instrument into a target language can be based on a variety of purposes. Casagrande (1954) has presented a review of four translation types, classified according to their general purpose. Pragmatic (accurate literal communication), aesthetic-poetic (expression of moods, feelings, and affect), linguistic (structural and grammatical equivalence), or ethnographic (conveyance of meaning and cultural content) translation can be the ultimate goal of the translation process. Yet within the content of attitudinal measurement, it is the desire to have equivalent conceptual measurement across subpopulations so that the same construct is being measured for each group. Therefore, an ethnographic type of translation was sought as the ${}_{I}P_{I}$ was translated from English to Japanese. The translators used for this study were familar with both the Engligh and Japanese cultures as well as versed in both languages. Some latitude was granted to the

translators because there were instances in which there were no cultural equivalent wordings or themes. The six scales of the $_{I}P_{I}$ were independently translated by three fluent bilingual individuals. The three translators then jointly resolved their differences and proposed a final version. To verify the quality of the translation, four Japanese researchers retranslated the instrument into English. Items that did not sufficiently converge to their original English meanings were revised.

The sample for this study was composed of 498 Japanese businessmen and educators and a comparable sample of 746 Americans. [1] The samples were essentially equivalent in age, 36 and 37 for Japan and America, respectively. The Japanese sample contained a greater percentage of males, 67% versus 57%. The American sample was the more educated of the two groups.

The $_{I}P_{I}$ was designed to provide equivalent measurement across various groups and organizations, and possibly cultures although this has never been directly tested. The basic hypothesis was that individuals from either culture with identical standing on any one of the six latent incentives scales would have identical expected observed scores on the same scale; i.e., an identical factor structure would exist across the groups.

Normality Assessment and Subscale Formation

The initial step in testing the equivalence of measurement for the $_{I}P_{I}$ was to determine the extent to which the inventory satisfied the requirement for normally distributed item responses. Therefore the correlation between each item response and its corresponding propensity variable, $\rho(X_i, L_i)$, was be determined for all forty-seven items (Drasgow & Dorans 1982). Dependent upon the obtained values, it might be necessary to form subscales of the items.

Item content was the most important criterion in the construction of the subscales. Given that the $_{I}P_{I}$ was composed of six relatively distinct

scales, subscales were created within each of these "content" areas. Thurstone (1947) stated that at least three elements (items) should represent each factor of an instrument. This resulted in the formation of at least eighteen subscales to represent the six scales of the ${}_{I}P_{I}$.

Subscales were used for all subsequent analyses regardless of whether or not the item responses were normally distributed. Such a strict adherence to the use of subscales as the unit of measurement was required even though item analysis was preferable. The use of subscales was necessary due to the fact that the ${}_{I}P_{I}$ was composed of so many items and the computional methods for determining maximum likelihood solutions were extremely time consuming when working with so many variables. If the subscale analyses showed that equivalence of measurement did not extend across the cultures then the analysis would be repeated for the biased subscale but on an item level (if the items themselves have been determined to be normally distributed).

Determination of the Optimal Factor Solution

Once the eighteen subscales were constructed for the six ${}_{I}P_{I}$ scales, it was necessary to conduct a preliminary factor analysis. Such an analysis of the subscales provided initial starting values for the maximum likelihood factor analysis. By inputting such values, the number of iterations required before convergence to the maximum of the likelihood function was greatly reduced. Based on the factor loadings that were calculated in this preliminary analysis, a maximum likelihood factor analysis was conducted using Joreskog and Sorbom's (1981) LISREL computer program to estimate all parameters for the maximum likelihood solutions. In order to obtain a unique factor solution from LISREL, it was necessary to constrain $\underline{m}^2$ parameters ($\underline{m}$ = number of common factors) of the factor model (Joreskog & Sorbom 1979). This was achieved by fixing the diagonal elements of ϕ at unity and $\underline{m}-1$ elements in each column of

Λ_y at zero. To determine which elements to fix at zero, the largest loading in each column of Λ_y was found. All elements of the $m-1$ corresponding columns were fixed at zero. The factor model was then estimated and its "goodness of fit" was determined.

Joreskog (1967) developed a large sample χ^2 test for assessing the fit of the maximum likelihood estimates to the observed data. A significant χ^2 is interpreted as implying that one or more common factors is required to explain the covariances of the observed variables (or possibly there is model misfit -- i.e., the variables are not multivariate normal). McGaw and Joreskog (1971) warn against the strict adherence to the χ^2 statistic because of its sample size dependency. Rather, Joreskog and Sorbom (1981) suggest that the χ^2 value be used as a measure of the goodness of fit for the hypothesized factor model. Large χ^2 values are indicative of bad fitting models and small χ^2 values of good fitting models. The degrees of freedom for the factor model can be used to gauge the magnitude of the χ^2 values. This approach is especially applicable to comparisons of different factor models by using a ratio of change in χ^2 to change in degrees of freedom (Kroonenberg & Lewis 1982).

Joreskog and Sorbom (1981) provide two other measures of the goodness of fit for the factor model. The first of these is called the goodness of fit index (GFI). The GFI is independent of sample size and is relatively robust to departures from normality by the observed variables. An alternative to the GFI that uses mean squares instead of total sum of squares in its calculations is the adjusted GFI. The AGFI takes the degrees of freedom for the factor model into account when computing the statistic. Both measures are an indication of the relative amount of variance and covariance jointly accounted for by the factor model and should range in value from zero to one. "Unfortunately, the statistical distribution (of GFI and AGFI) is unknown, even under idealized assumptions, there is no standard to compare it with" (Joreskog & Sorbom 1981,

p. I.41). The second measure of goodness of fit is the root mean square residual (RMR), an index of the average of the residual variances and covariances. The value of RMR can only be interpreted with respect to the observed variances and covariances, and generally the smaller this value, the better the fit.

While each of these statistics provide pertinent information concerning the overall fit of the factor model, in no way do they reflect the quality of the factor model (rather, only an assessment of the overall fit of the model to the sample data). It is possible even if the indices show that the overall fit is good that the underlying relationships of the model may be poorly determined. To detect such a possibility it is important to examine the residual variances and covariances ($\underset{\sim}{S} - \hat{\underset{\sim}{\Sigma}}$). Since $\underset{\sim}{S}$ is not always expressed as a correlation matrix, Joreskog and Sorbom (1981) calculate normalized residuals. This term is approximately a standard normal variable although it is correlated from cell to cell. The normalized residual serves as a flag of insufficient fit by the model if its value is larger than two for a particular element of the $\underset{\sim}{S}$ matrix. Specific corrections of either relaxing or constraining additional parameters of the model can be implemented according to "modification indices" (Joreskog & Sorbom 1981). The index is associated with the derivatives of the fitting function for all fixed and constrained parameters and is distributed as a χ^2 with one degree of freedom. If the fixed parameter with the largest modification index is freed, a decrement of <u>at least</u> the magnitude of the index will result in the associated χ^2. The process follows such that successive parameters with the largest index are freed one at a time until the change associated with χ^2 is no longer unacceptable.

The maximum likelihood factor solution was further refined in order to minimize overfitting of the factor model by fixing additional parameters of

Λ_y at zero (Kroonenberg & Lewis 1982). One of the benefits of maximum likelihood estimation is that standard errors can be calculated for each parameter of the factor model. From the standard errors, a t-statistic can be computed for each element of Λ_y to determine whether it was statistically different from zero. Those t-statistics small in magnitude were likely candidates to be fixed at zero. Once these elements had been identified, the LISREL analysis was repeated to determine if the additional constraints seriously affected the fit of the factor model.

The χ^2 value for the simplified model (more parameter fixed at zero) should not substantially exceed the χ^2 for the original model given the number of imposed restrictions. If the ratio of change in χ^2 to the change in degrees of freedom was large, the additional constraints imposed on the factor model caused a departure by the factor model from the observed data. Therefore the number of constraints (number of parameters fixed at zero) must be relaxed. Using the modification indices, the exact misfitting parameter can be identified and accordingly be freed to provide a better fitting model. In summary, the factor model for which the parameters with t-statistics less than 2.58 (alpha = .01) in absolute value were fixed at zero served as the upper limit for the "optimal" factor model. Further refinements were made on the model by freeing the parameter with the largest modification index (only one parameter was freed at a time). This process was continued until subsequent freeing of parameters did not significantly decrease the χ^2. The resulting factor model would serve as the foundation for the analysis of the American versus the Japanese cultures.

SIFASP for the Cultures

The SIFASP analysis was a process of sequential hypothesis testing. Dependent upon the conclusions of the significance tests, the analysis either

branched to another test or terminated. This progression of significance tests is given in Figure 1. The figure provides an overview for the SIFASP analysis and will serve as a useful reference for the following discussion. The null hypothesis for each test was given along with the test statistic that was used to determine the conclusion for that analysis. The course of action was given for both significant and nonsignificant results.

Before the SIFASP was conducted, it was necessary to rescale the eighteen subscales since their units of measurement vary across subscales due to differing numbers of items for each. Typically the variance-covariance matrices for the subpopulations are rescaled to correlation matrices, yet such a transformation would eliminate important group differences (McGaw & Joreskog 1971). An acceptable rescaling technique was based on a weighted pooled variance-covariance matrix, $\underset{\sim}{S}$:

$$(2) \quad \underset{\sim}{S} = \frac{1}{N - g} \sum_{i=1}^{g} (N_i - 1)\, \underset{\sim}{S}_i \, .$$

where $N = \sum N_i$ and $i = (1, 2, \ldots, k)$, the number of subpopulations.

A diagonal matrix of the reciprocals of subscale standard deviations was computed from $\underset{\sim}{S}$:

$$(3) \quad \underset{\sim}{D} = (\text{Diag}\, \underset{\sim}{S})^{-1/2} \, .$$

The original variance-covariance matrices was then rescaled by this diagonal matrix:

$$(4) \quad \underset{\sim}{S}_i^{*} = \underset{\sim}{D}\, \underset{\sim}{S}_i\, \underset{\sim}{D} \, .$$

The preliminary test in a SIFASP analysis was to determine if the rescaled variance-covariance matrices for the cultures were equivalent. If the matrices were not significantly different, then equivalence of measurement held across the groups and the analysis stopped. Box (1949) developed a F test to test whether there was equality across these matrices. If the null hypothesis of equality was rejected, subsequent tests were performed to test the invariance property of Λ_y.

Dependent upon the significance of the F-statistic for the above test, a separate maximum likelihood factor solution would be computed for each culture. These independent solutions followed the simplified optimal factor solution that was calculated for the entire sample; i.e., the same number of common factors was used and the same elements of Λ_y were held constant, fixed at zero, or set free to be estimated. The strategy was to compare this "unconstrained" factor model with a model in which Λ_y and Θ^2 were required to identical across cultures.

Since the cultures were mutually exclusive and independent, the respective χ^2 values associated with the separate maximum likelihood solutions were also independent. The sum of the independent χ^2 values was also distributed as a χ^2 and would accordingly represent the χ^2 that resulted from letting Λ_y, Φ, and Θ^2 vary across cultures. The χ^2 value for the "constrained" model was determined by a simultaneous factor analysis of each rescaled variance-covariance matrices, S_i^*. In this analysis, Λ_y and Θ^2 were required to be the same for both cultures while the entire Φ matrix was set free to vary. By totally freeing Φ, an additional m constraints must be imposed on the Λ_y matrix to ensure that the maximum likelihood solution was unique (Joreskog & Sorbom 1979). Thus an additional non-zero element (the largest loading) was fixed in each column of Λ_y to account for the m constraints lost by freeing the diagonal of Φ. The difference in the χ^2 values for the two

models was tested to determine if the requirement of invariance for Λ_y and Θ^2 significantly affected the fit of the unconstrained factor model. If the resulting ratio of χ^2 values for the test of American versus Japanese samples was significant, the conclusion would be that constraining both Λ_y and Θ^2 across the cultures caused an inordinate degree of variance in the factor model and thereby causing it to not sufficiently fit the observed data.

A follow-up analysis in which Θ^2 was not constrained to be equal across both groups would test the invariance property of Λ_y. If the resulting ratio of χ^2 to degrees of freedom was small then the conclusion would be that only the uniquenesses, Θ^2, were not invariant across the groups and equivalence of measurement held. However, if the ratio were small, Λ_y would not be invariant across the cultures and there would be no equivalence of measurement. The task then became one of identifying the exact subscale or subscales which were the locus of effect and repeating the analysis but on an item level.

Testing the Equality of Subscale Factor Loadings Across Cultures

The primary concern before conducting the test of equality of subscale factor loadings was that the estimated loadings be expressed in comparable metrics (Mulaik 1972). Such caution was essential to ensure that consistent differences among the estimated loadings were not deemed as being significantly different merely due to the fact that the loadings were based on dissimilar scales of measurement. To achieve comparable metrics for the estimated loadings, the variance-covariance matrix for each subpopulation was rescaled as was done earlier according to equations 2-4. These rescaled variance-covariance matrices then served as input for the factor analyses as opposed to using the original correlation matrices. In addition to rescaling the variance-covariance matrices, it was necessary to also express the factor variances in similar units. Drasgow and Kanfer (1983) suggested that the

variances (the diagonals of the Φ matrices) be fixed at the values that were obtained in the earlier simultaneous factor analysis for which the Ψ^2 and Φ matrices were entirely free and Λ_y was held invariant across the groups. Having completed both of these steps, the estimated loadings resulting from the separate maximum likelihood factor analyses were essentially in analogous metrics and would be directly comparable.

The test statistic for such comparisons was developed by Drasgow and Kanfer (1983) and followed a χ^2 distribution with its degrees of freedom equal to the number of estimated loadings (p) for a particular subscale. The comparison test was analogous to the univariate test for difference in two means, but extended to the multivariate case (see Tatsuoka 1971). First, a vector of differences between the estimated factor loadings of two subpopulations was computed. These differences followed a multivariate normal distribution with each element of the vector having an expected value of zero under the null hypothesis. Second, the variance-covariance matrix of the differences was formed by adding the variance-covariance matrices of the two cultures.

Factor loadings computed on the basis of maximum likelihood estimation procedures have asymptotic multivariate normal sampling distributions. Therefore in large samples, it was possible to determine the sampling variance and covariances for each loading that was estimated. The sampling variances were calculated by squaring each standard error of the estimated loadings. Likewise, the sampling covariances were the products of all pairwise standard errors multiplied by the accompanying correlation between the two loading estimates. Sampling variances and covariances were determined for each subpopulation:

$$(5) \quad \underset{\sim}{S}_i = \begin{bmatrix} SE_{i1}^2 & & & \\ SE_{i1}\,SE_{i2}\,r_{12} & SE_{i2}^2 & & \\ \vdots & \vdots & \ddots & \\ SE_{i1}\,SE_{ip}\,r_{1p} & SE_{i2}\,SE_{ip}\,r_{2p} & \cdots & SE_{ip}^2 \end{bmatrix}$$

The variance-covariance matrix of the differences ($\underset{\sim}{A}$) was computed by summing the variance-covariance matrices for each culture; i.e., $\underset{\sim}{S}_1$ and $\underset{\sim}{S}_2$. Finally, the test statistic was calculated by pre- and post-multiplying the inverse of the variance-covariance of the differences by the vector of differences ($\underset{\sim}{d}_i$):

$$(6) \quad \chi^2_i = \underset{\sim}{d}_i' \; \underset{\sim}{A}^{-1} \; \underset{\sim}{d}_i$$

with $\underline{p}$ degrees of freedom.

These comparisons identified those subscales which had inconsistent factor loadings across the cultures.

Analysis of Subscales Identified as Being Inconsistent

The subscales that were identified as having inconsistent factor loadings across the cultures were then expanded to their original items. In this manner, the exact item or items which were the cause of the overall inconsistency could be determined. It is important to recall that the item responses need to be approximately normally distributed to justify such an analysis. If the item responses were not sufficiently normally distributed, conclusions could only be based on the subscales analyses. Given that the responses were approximately normal, the procedure for the analysis of the

items was essentially the same as that for the analysis of the subscales.

For the ${}_{I}P_{I}$, the subscales would be formed of two to four items. As a result of expanding the inconsistent subscales to the item level, the number of possible variables entering into the analysis might become too large for the LISREL program (recall that subscales were originally used in order to reduce the number of variables to a manageable level). So that the item analysis would not be overwhelming, the only variables entering into the item analyses were the original items from the inconsistent subscales and a reference subscale for each inconsistent subscale.

As stated earlier, the analysis of the items and the appropriate subscales basically followed the procedures discussed for the SIFASP. First, a continuity assessment based on $\rho(X_i, L_i)$ for each item of the inconsistent subscale was conducted to determine if an item analysis was justified. Assuming that the items were sufficiently normal, the subscales for inclusion in the analysis were then identified. Separate maximum likelihood factor solutions based on the correlation matrices of items-subscales for both cultures were computed. The starting values for the Λ_y, Φ, and Ψ^2 matrices were the LISREL estimates determined in the previous maximum likelihood factor solutions for each culture. For Λ_y and Ψ^2, modifications were made of the LISREL estimate matrices to only include the appropriate subscales. The starting values for the items forming the inconsistent subscale were the LISREL estimate for that subscale repeated for each item. The resulting χ^2 value for each of the solutions was then summed and represented the overall χ^2 for the unconstrained factor model in which Λ_y, Ψ^2, and Φ were free to vary across subpopulations.

It was necessary to compute rescaled variance-covariance matrices for both cultures to use as input for the overall SIFASP. These matrices were rescaled according to equations 2-4 and included only the items and the appropriate

subscales. In addition to rescaling these matrices, the diagonal elements of the Φ matrix were set free to be estimated. This required that an additional m elements of the Λ_y matrix be fixed. The starting values for the SIFASP were taken from the previous SIFASP analysis in which Λ_y was held invariant. Again, modifications of the LISREL estimates were made so that only the appropriate subscales were included and that the starting values for the items reflected the estimates for the inconsistent subscale. The χ^2 resulting from the SIFASP represented the degree of fit for the constrained factor model in which Λ_y was held invariant across groups. This χ^2 was then compared with the χ^2 for the unconstrained factor model. If the increase was significant, a test would be conducted to determine the invariance property of Λ_y for each item. It was almost certain that this test would be significant since it was based solely on the items of previously determined inconsistent subscales.

Having confirmed the previous analyses that identified potentially inconsistent subscales, the task was to identify those items that were inconsistent across the groups. Recall that it was essential that the factor loadings were expressed in comparable metrics. A common metric for the estimated loadings was achieved by first using rescaled variance-covariance matrices and then by fixing the factor variances at the values obtained in the earlier SIFASP for which Ψ^2 and Φ were free and Λ_y was held invariant. The χ^2 statistic was then calculated for each item (not each subscale) that formed the inconsistent subscale. For those items in which the statistic was significant, the conclusion was that the estimated factor loadings were not consistent over the groups.

It is necessary to digress at this point to reflect on the number of statistical tests that were being made for a particular analysis and the accompanying possibility of committing type I errors. In testing the equality

of subscale factor loadings across subpopulations, a test statistic was calculated for each subscale. In conducting so many tests, the likelihood of identifying a subscale as inconsistent when in fact it was not, had increased. Therefore the χ^2 statistic served as a liberal statistic in that the possibility of overidentifying the true number of inconsistent subscales definitely existed. At this point, the object was merely to identify the possible inconsistent subscales, so committing overidentification errors was preferable to underidentification errors. The alpha level was not corrected for the number of tests. Thus the χ^2 statistic served as a preliminary test to identify inconsistent subscales. As stated earlier, the SIFASP of the expanded inconsistent subscale and the appropriate subscales results in a confirmation of the χ^2 statistic. It was because of this confirmatory analysis that such liberal precautions were taken in the previous analysis. Having identified the specific items of the inconsistent subscale for which the factor loadings were not equivalent across cultures, only a single pairwise comparison was made of the estimated loadings across the groups so no correction was made on alpha.

RESULTS

Subscale Formation and Preliminary Analyses

Eighteen subscales were formed from the forty-seven items of the ${}_{I}P_{I}$ on the criterion of item content. The correlations between the item responses and the corresponding propensity variables, $\rho(X_i, L_i)$, for the subscales showed a marked increase in the responses more nearly approaching normality, as opposed to the distributional characteristics of the individual items. While the item correlations ranged in value from .85 to .95, the creation of subscales improved the correlations to a range of .94 to .99 (see Table 1).

Thus the ${}_{I}P_{I}$ did yield item responses that were sufficiently normal and accordingly the items could be analyzed in their own right. But the analysis centered on subscales for reasons discussed earlier and the possibility for regressing back to the item level was definitely warranted.

Having fixed the diagonal elements of Φ and $m-1$ elements in each column of Λ_y to insure a unique solution, the maximum likelihood procedure used the starting values of the preliminary factor analysis. This factor solution was further analyzed to determine the nonsignificant factor loadings. Since LISREL provides t-statistics for parameters that were estimated, a series of analyses was conducted which fixed at zero all estimates less than a certain t-value. Table 2 shows the results from the analyses in which the t-value was incremented by 1.0 for each factor solution. The model with the minimum number of fixed parameters was the one with m^2 or 36 parameters fixed. This model with 60 df and a χ^2 of 378.34 served as the comparison model against which all changes in df and χ^2 were determined. Using the ratio of change in χ^2 to change in degrees of freedom as an indication of fit between the factor model and the observed data, the analysis for an upper limit of t equal to 3.0 resulted in a ratio of 3.11. While this was a moderately large ratio, this analysis did not greatly differ on the other goodness of fit indices for the analysis of t equal 2.0.

To improve the fit of the model at the 3.0 t level, the modification indices were computed for the 79 fixed parameters. This index served as a means of identifying parameters that would significantly increase the fit of the model if relaxed (not constrained to equal zero). Repeating the analysis but with the identified parameter not fixed resulted in a ratio of 2.63, see the refinement section of Table 2. This process of determining the largest modification index, freeing the parameter, and calculating the χ^2 associated with the new model was repeated until the decrease in the χ^2 between models

was no longer significant ($\chi^2_1 = 6.64$ for alpha = .01). The model with 74 parameter fixed satisfied this criteria in that freeing an additional parameter only resulted in a χ^2 change of 5.32. The ratio of changes in χ^2 to df associated with this model was 1.87, a marked improvement over the 3.11 value for the 79 fixed parameter model. Table 3 presents the estimated factor loadings and uniquenesses for the optimal factor solution. There was a clear and compelling simple structure for the total sample of 1244 subjects. Table 4 presents the estimated factor variance-covariance matrix. Since the variances were fixed at unity, the matrix can be interpreted as the factor correlation matrix. These three matrices served as the foundation for all subsequent analyses.

SIFASP Analysis of Subscales

Because the units of measurement were arbitary for the subscales, the first step in the SIFASP was to rescale the variance-covariance matrices for each group according to equations 2-4. Box's test for equality of the two matrices resulted in a significant F ($p < .001$) implying that the matrices differed. Since the null hypothesis of equal matrices was rejected, a SIFASP was conducted to determine which component of the variance-covariance matrices was the cause of the discrepancy.

Separate maximum likelihood factor solutions for the subscales were obtained for each subpopulation. Each analysis was based on the optimal factor solution with the values in Tables 3 and 4 serving as the starting values. The parameters fixed at zero, one, or a constant were still fixed in these analyses. It was necessary to determine whether the common elements fixed at zero across the groups were actually zero in both cultures. The SIFASP analyses might not be able to detect factor loadings that were nonzero in one culture while they were fixed at zero. To determine if the fixed parameters

had any effect on the analysis, the modification indices for the two cultures were examined for significance. Only one of the parameters resulted in significant modication indices ($.01 < p < .05$) for the two groups. The parameter was from the American sample, although all of the goodness of fit indices indicated that the factor model better fit the American data than it did the Japanese data, see Table 5. Since there was no inordinate difference in the goodness of fit indices, the analyses continued as before with the same parameters fixed across both groups.

The χ^2 for the maximum likelihood factor solutions for both cultures is shown in Table 5. Since the groups were independent, the associated χ^2 values were independent and their sum (407.10) represented a χ^2 with 196 df. This χ^2 resulted from letting $\underset{\sim}{\Lambda}_y$, $\underset{\sim}{\Phi}$, and $\underset{\sim}{\Theta}^2$ vary across subpopulations, and is referred to as the unconstrained model. Note the magnitude of the ratio of χ^2 to df -- 2.08, this will serve as a reference with which to compare the resulting ratios of the other SIFASP analyses. Table 6 presents the factor loadings matrices resulting from these separate factor solutions.

Next, a simultaneous factor analysis was conducted in which both $\underset{\sim}{\Lambda}_y$ and $\underset{\sim}{\Theta}^2$ were constrained to be identical across the two cultures. The $\underset{\sim}{\Phi}$ matrix was completely free to be estimated and thus an additional $\underline{m}$ nonzero elements in $\underset{\sim}{\Lambda}_y$ were fixed. Fixed and free parameters (except for the above mentioned changes) were set according to the values in Tables 3 and 4. The χ^2 for this constrained model was 570.16 for 242 df, an increase in the χ^2 of 163.06 for additional 46 df. This increase in χ^2 was significant ($p < .01$) but the ratio of change in χ^2 to change in df was relatively moderate, 3.54.

A second simultaneous factor anlaysis was conducted but with only $\underset{\sim}{\Lambda}_y$ constrained across the groups. Both $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Theta}^2$ were free to vary across the subpopulations. Imposing the restraint of equal factor loadings across the three groups resulted in an increase of 93.99 for χ^2 and 33 additional df,

suggesting that a different Λ_y matrix for each subpopulation better approximated the data than did some overall Λ_y matrix for both groups. Both the GFI and RMR indices indicated an improvement in the quality of fit as Θ^2 was freed to vary across subpopulations. Of course, the model in which only Λ_y was invariant showed better fit because this was more general (less parameters was constrained) than the model in which both Λ_y and Θ^2 were constrained. Thus it was necessary to compare the models with respect to their number of constrained parameters as reflected in the change in df. A ratio of change in χ^2 to change in df was an indication of model fit across models. Although the increase in χ^2 was significant for constraining Λ_y, the ratio of changes, 2.84, was relatively small compared to the magnitude of the ratio for the unconstrained model (2.08).

Finally, a simultaneous factor analysis was conducted in which only Θ^2 was invariant across the subpopulations. The increase in χ^2 (63.24) was significant ($p < .01$) for its 16 df. Again, GFI and RMR indicated this model as better fitting the data because there were less fixed parameters. Judged with respect to the change in df, the model yielded a ratio that was large (3.96). Thus there was some evidence that the uniquenesses were <u>not</u> invariant across subpopulations. The results from this simultaneous analysis and the previous analysis suggested that the violations in equivalence of measurement across cultures were more the result of differences among the unique variances than in differences in the factor pattern matrices. Although this was the case, the factor loadings were further analyzed to determine the subscales and subsequently the items of th. $_I P_I$ which had inconsistent factor loadings across the groups.

In identifying the inconsistent subscales, the estimated loadings were expressed in comparable metrics by using rescaled variance-covariance matrices and fixing the factor variances. The factor variances ranged from .872 to

1.178 for the American culture and from .853 to 1.032 for the Japanese culture. Computing the vector of differences and variance-covariance for these differences, the χ^2 statistic identified four inconsistent subscales ($p < .01$): subscales 2, 5, 12, and 17. These subscales were expanded to their original items (ten items composed the four subscales) for further analysis.

SIFASP Analysis of Items

The correlations between the item responses and the corresponding propensity variables, $\rho(X_i, L_i)$, ranged in value from .87 to .96 for the ten items that formed the inconsistent subscales. Such values were sufficiently high to warrant the analysis being extended to the item level. In addition to these ten items, six other subscales were included in the analysis to serve as reference variables and to insure that at least three variables define each factor.

The procedure deviated at this point from the previous subscale analyses in that the SIFASP analysis was conducted first. The SIFASP analysis provided the values at which to fix the factor variances in order to obtain a proper scaling of the factor loadings as well as serve as a confirmation of the χ^2 tests for the subscales. The factor variances determined in the SIFASP analysis in which Λ_y was invariant but Ψ^2 free, ranged in value from .876 to 1.029 for the American sample and .722 to 1.016 for the Japanese sample. Fixing the diagonal of each Φ matrix at these determined values as well as using the rescaled variance-covariance matrices for each group ensured that the estimated factor loadings were comparable. The starting values for the Λ_y, Ψ^2, and Φ matrices in the separate maximum likelihood solutions were provided by the estimates generated in the SIFASP analysis that identified the inconsistent subscales. The subscale estimate was repeated for each item of the inconsistent subscale. The estimates of the factor loadings of the ten

items and their standard errors are presented in Table 7 for each subpopulation. The standard errors were relatively consistent across the groups, generally ranging in value from .03 to .09.

The χ^2 statistic was calculated for each of the ten items and the significance levels are presented in Table 8. The pairwise comparisons of loadings classified six items as inconsistent for alpha < .05. The other four items were not classified as inconsistent by the pairwise comparisons. There was no consistent pattern in the item response-item propensity correlations, both high and low values occurred for the inconsistent items.

DISCUSSION

The analysis of the translation of the ${}_{I}P_{I}$ by means of maximum likelihood factor analysis found six of its forty-seven items to provide nonequivalent measurement across the American and Japanese cultures. Such inconsistent measurement can be the result of: (1)inadequate translation of the instrument from the base to the target language, (2)actual differences in the conceptual traits across the two cultures (i.e., differences in cultural perceptions of the traits being measured), or (3)errors in or insensitivity of the factor analytic detection procedure in distinguishing between consistent and inconsistent items. Surely some combination of these three causes contributed to the observed differences in the factor structures of the two cultures. The following discussion is divived into two sections. First, the inconsistent items will be intrepreted in terms of item content, format, adequacy of translation, and cross-cultural meaningfulness. Second, the maximum likelihood factor analysis procedure will be evaluated with respect to identifying inconsistent items in the context of assessing attitudinal scale translations.

Interpretation of Inconsistent Items

The six inconsistent items identified by the pairwise comparisons were examined for consistencies of content, format, translation, and meaning. The six items are presented in Figure 2 as well as the $_{I}P_{I}$ scales to which these items belong. Note that two inconsistent items are from both the task involvement and financial rewards scales while only one item each is from the competition and social concern scales.

Item	$_{I}P_{I}$ Subscale
Were totally responsible for the outcome.	Task Involvement
Find work fun.	Task Involvement
Need to be number one.	Competition
Commitment to social and civic causes.	Social Concern
I work hard for salary increases and promotions.	Financial Rewards
When personal payoff is involved, I work harder.	Financial Rewards

Figure 2. Inconsistent items for the American and Japanese cultures.

With respect to just item content, the "totally responsible" theme of the first item is unique to the task involvement scale. The other items of the scale reflect more the intrinsic nature of personal involvement and not the sense of duty implied by responsibility. For the other inconsistent item of the scale, "find work fun," the idea of "fun" is again not found in the other items. The inconsistent items identified for the competition and social

concern scales do not readily characterize themes that are incongruous with the other items of the scales. The two inconsistent items of the financial rewards scale are ambigous with respect to other items of the scale in terms of "personal payoff" and "promotion." All other items specifically use the terms money, wealth, or monetary.

It is important to note the two different formats of the inconsistent items. Half of the $_{I}P_{I}$ items had a format like that of the items for the financial rewards scale: a statement which required a one to five response on a scale of strongly disagree to strongly agree. The other format, like that of the first four inconsistent items, required the respondent to project his/her perceptions of successful people for a list of personal characteristics. The fact that more items were classified as inconsistent for this second format may reflect that the use of these formats are inconsistent across the cultures or that the instructions to these sections were not adequately translated. Further investigation is needed in this regard to determine the possible influence of varying formats. While themes within the $_{I}P_{I}$ scales can be identified for the inconsistent items and possible influence of changing formats noted, it is more informative to address the issue of proper translation and cross-cultural meaning of the constructs within the context of scale translations.

To address the question of scale translation, another group of five translators (different from the first set) were asked to review the inconsistent items. They first translated the Japanese into English unaware of the true original English version. This blind rendition served as the basis for which they assessed the fidelity of the Japanese translation as well as noted awkward phrases or wording. Being familar with both cultures, the translators were asked to examine the items for possible cultural differences that may have lead to the inconsistent measurement.

The first item, "were totally responsible for the outcome," was thought to be adequately translated but there was some question about the responsibility - outcome relation. In Japanese, the item has two connotations -- being morally responsible or being the actual agent to whom the outcome is attributed. It is this second idea that the English version seeks to address, not the deep-rooted virtuous issue of the first. This difference in connotations of the item could possibly contribute to its inconsistent measurement. The opportunity to be totally responsible for an outcome is much less likely within the Japanese work setting than it is in America. Employees are simultaneously a member of as many as eight to ten work groups, each with a different task. Ouchi (1981, p. 41) noted in an interview with a Japanese manager that "no one can come up with a work improvement idea alone. We work together, and any ideas that one of us may have are actually developed by watching others and talking to others. If one of us was singled out for being responsible for such an idea, it would embarass all of us." Thus there is the possiblity that a Japanese individual would have a different perception of this item than would an American and this may influence the factor structures for each group.

It was the concensus of the translators that the meaning of the second item, "find work fun," was not sufficiently conveyed in the Japanese translation. The difficulty lies in the translation of the word "find." The English version implies a much more sophisticated concept of positive motivation resulting from work while the Japanese version is a bit less definitive. The organizational structures significantly differ across the two cultures in that Japan emphasizes non-specialized career paths while the exact opposite is true for American organizations. Hackman and Oldman (1976) have shown that workers who are continually faced with new activities within their jobs are more productive and satisfied than those who have only specialized responsibilities. Such differences in the design of the organizational

structure may have contributed to differing perceptions of this item.

Again, the translators were in total agreement with respect to the "need to be number one" item. The word "need" was translated as "want" in the Japanese version which clearly is a less emphatic expression, not expressing the ambitous and competitive nature of the English original. Therefore, there is definite problems in the translation of this item which may have significantly influenced the factor structure across the cultures. Yet again a strong case can be made for cultural differences on this item. Lifetime employment is much more prevalent in Japan although this trend has shifted in recent years (Hayaski, 1982). Such employment policies create stability within the organization and discourages any one to strive for immediate self-fulfilling goals at the expense of other individuals (Christopher 1983). Coupled with the non-specialization career orientation of the Japanese, there is little incentive among individuals to be "number one." Such non-competitiveness among individuals within the Japanese organizations may possibly influence their responses to this item and accordingly account for the nonequivalent measurement across cultures.

The next item, "commitment to social and civic causes," also had problems in its translation to Japanese. The Japanese version reflects mere "participation," a much lower order of involvement than commitment which does not entail a sense of obligation or noblesse oblige. Obviously, one can participate in a cause more casually than if one is committed to it. Also the Japanese word for "cause" was translated as "activity," again not reflecting a sense of obligation. "Social" had to be translated as "public" because in Japan it carries a negative connotation associated with socialism or the extreme left.

The last two items from the financial rewards scale were considered to be faithful translations of the English original. A few translators however did

comment on the differing employment systems of the two cultures. It is important to recognize that the Japanese system is generally more strongly time-in-grade and seniority-based for compensation and promotion than is the American system (Christopher 1983). This is not to say that there is no merit-based differentation in Japan but rather the emphasis is clearly a more long range perspective. Ouchi (1981) develops a scenario of a newly hired college graduate, Sugao, to illustrate the promotion and evaluation practices of a typical Japanese organization: "For ten years, Sugao will receive exactly the same increases in pay and exactly the same promotions as the other fifteen young men who have entered with him. Only after ten years will anyone make a formal evaluation of Sugao or his peers; not until then will one person receive a larger promotion than another" (p. 22). This does not imply that employees do not work hard, it is that their motives are not based on personal financial gains but rather on the prosperiety of the organization. Such differences in corporate ideology could have influenced the factor structure of the two cultures.

Surely some combination or interaction of the quality of translation, cross-cultural differences, and possibly item format that ultimately caused the differences in the observed factor structure. It appears that inadequate translation was the culprit for three items: "find word fun," "need to be number one," and "commitment to social and civic causes." Cultural differences in perceptions of the items may have caused the nonequivalent measurement for the other three items.

Assessment of the Maximum Likelihood Factor Analysis Procedure for the Detection of Item Inconsistencies Across Cultures

The maximum likelihood factor analysis procedure described and implemented in this study provided a particularly accomodating and statistically rigorous

means of detecting item inconsistencies in the translation of an attitudinal instruments. This section will evaluate the procedure as it was applied to the $_{I}P_{I}$ for the purpose of assessing the equivalency of measurement across both cultures. The $_{I}P_{I}$ is representative of most attitudinal measures in that it is distinctly multidimensional, only moderate in length, and possesses a fixed number of response alternatives. In this regard, the generalizations resulting from this study concerning the use of maximum likelihood factor analysis should extend to other attitudinal instruments. The following discussion is basically a step by step commentary of the maximum likelihood procedure, highlighting both its strengths and weaknessess.

The selection of individuals from both cultures was determined on the basis of similarity of profession (i.e., business men and educators). Although the groups may possibly have been heterogenous within their respective cultures, if larger samples were available, comparisons could have been made on a more refined level between Japanese and American business men or educators. Despite the lack of such pure samples, the procedure was still able to function with subpopulations numbering only about 500 subjects. In fact, LOGIST (Wood, Wingersky, and Lord, 1976) did not converge after fifteen iterations even though the entire sample of 1244 subjects was used. Whether this was the result of "small" sample size, the violation of the unidimensionality assumption, or some combination of the two, it further reinforced the need for a multidimensional analysis of our attitudinal measure.

An important limitation of the study involved the normality requirements for the factor model -- the subscales and especially the items were neither normally distributed nor continuous. Although the $\rho(X_i, L_i)$ values were relatively high for the subscales and acceptable for the items, this was not a definite indication that the variables were normally distributed. Such a violation of the assumption for maximum likelihood estimation may result in

distorted estimates for the components of the factor model. Accordingly, if the factor model was poorly estimated the test statistics were inflated and were no longer an accura'e indication of the goodness of fit by the factor model.

All analyses of the subscales, determining the optimal factor solution and identifying the inconsistent subscales by the SIFASP analyses, quickly identified unique factor solutions, usually within four to six iterations. Such alacrity in achieving the solution was likely the result of extremely good starting values from the preliminary factor analysis and high $\rho(X_i, L_i)$ values for all the subscales. While such quality in estimation of the solutions was evident, this did not address the concern of the quality of fit by the factor model. The residuals, normalized residual, and the modification indices of the subscales all indicated that the factor model more than adequately fit the data. In summary, the maximum likelihood procedure readily accomodated the analyses of subscales.

The quality of estimation and fit by the factor model was only slightly less for the item level analyses. Estimation of factor model parameters for the items took more iterations before convergence was reached, generally eight to sixteen. The additional iterations may have been required for a variety of reasons. First, the distribution of the item responses were less likely to approach normality. Second, the estimates for the inconsistent subscales were merely repeated for all items as its starting value, maybe these values were not realistic for all items forming the subscales. Third, the same parameters that were fixed at zero for the subscales analyses were again fixed on the item level. It was possible that the parameters for the items were not consistent across the groups as the subscales parameters were. The modification indices for the two cultures indicated that the fit of the model for the items would be slightly increased if a few parameters were set free.

Finally, the large number of statistical tests being made to determine inconsistent subscales and inconsistent items must be taken in account. Such testing may possibly lead to a number of significant results occurring by chance. In the initial analyses, steps were taken to minimize the number of "true" inconsistent subscales being misclassified by overidentifying the number of subscales considered inconsistent. Such precaution may result in an inordinate number of items being classified as inconsistent due to this overidentification of subscales.

In conclusion, the results in this study showed that inconsistent measurement can be identified in the analysis of cross-cultural scale transations for attitudinal instruments. The maximum likelihood factor analysis procedure provided many advantages over the classical test theory and item response theory approaches for the detection of inconsistent items in attitudinal instruments. Despite its limitations, the technique was readily able to accomodate multidimensional data sets that were polychotomously scored. Also the procedure provided a very rigorous statistical test that is deeply rooted in factor analysis research and is a viable alternative to the more restrictive item response theory techniques.

(1) Test the equality of variance-covariance matrices across subpopulations.

H_o: $\hat{\Sigma}_1 = \hat{\Sigma}_2$

Test statistic = Box's F.

	Significant	Nonsignificant
Conclusion	Matrices differ.	Equivalence of measurement.
Action	Test invariance properties of Λ_y & ρ^2.	End analysis.

(2) Test effect of constraining Λ_y and ρ^2 to being equal across subpopulations.

H_o: $\Lambda_{y1} = \Lambda_{y2}$ and $\rho_1^2 = \rho_2^2$

$$\text{Test statistic} = \frac{\chi^2_{\Lambda_y \rho^2} - \chi^2_{\text{unconstrained}}}{df_{\Lambda_y \rho^2} - df_{\text{unconstrained}}} .$$

	Significant	Nonsignificant
Conclusion	Λ_y or ρ^2 is not invariant.	Equivalence of measurement.
Action	Free ρ^2 and retest.	End analysis.

(3) Test effect of constraining Λ_y to be equal across subpopulations.

H_o: $\Lambda_{y1} = \Lambda_{y2}$

$$\text{Test statistics} = \frac{\chi^2_{\Lambda_y} - \chi^2_{\text{unconstrained}}}{df_{\Lambda_y} - df_{\text{unconstrained}}} .$$

	Significant	Nonsignificant
Conclusion	Λ_y is not invariant, equivalence of measurement does not hold.	Equivalence of measurement but ρ^2 is not invariant.
Action	Identify subscales across which Λ_y is invariant. Repeat analysis but on item level.	End analysis.

Figure 1. Progression of SIFASP analysis.

Table 1

Stem and Leaf Diagrams for $\rho(X_i, L_i)$ for
Both Items and Subscales (Significant to Three Digits)

ITEMS

```
.96
.95  3 3 7 5 7 5 3 6 1 7 5 4
.94  8 7 2 1 9 2 1 2 7 1
.93  5 9 7 2 1 8 9 3 7 6 8
.92  6 9
.91  7 6 4
.90  5
.89  4 3 3
.88  9 6
.87  4 3
.88
.85  2
.84
```

SUBSCALES

```
.99  4 1
.98  7 1 1 3 9 3 5 8 9
.97  1 7 9 4 9
.96  8
.95
.94  8
.93
```

Table 2

Determination of Optimal Six Factor Solution for Eighteen Subscales

Number of Parameters Fixed	Upper t-value Limit	df	χ^2	GFI	AGFI	RMR	Change in df	Change in χ^2	Ratio
36	-	60	378.34	.958	.881	.022	-	-	-
51	1.0	75	383.90	.957	.902	.022	15	5.56	.371
72	2.0	96	455.87	.939	.892	.029	36	77.53	2.15
79	3.0	103	512.17	.932	.886	.033	43	133.83	3.11
Refinement Stage *									
78	-	102	488.80	.939	.898	.031	42	110.46	2.63
77	-	101	477.86	.941	.900	.030	41	99.52	2.43
76	-	100	468.93	.942	.901	.028	40	90.59	2.26
75	-	99	460.01	.943	.902	.027	39	81.67	2.09
74	-	98	449.53	.945	.904	.027	38	71.19	1.87
73	-	97	444.21	.946	.905	.027	37	65.87	1.78

* Refinements are made according to the largest modification index for the model yielding changes greater than 6.64 (alpha = .01, 1 df).

Table 3

Estimated Factor Loadings and Uniqueness for the IPI Subscales
Optimal Factor Solution Determined by Maximum Likelihood Factor Analysis

	Factor						
Subscale	I	II	III	IV	V	VI	h^2
1				.83 (03)			.32 (03)
2		.08 (02)	.10 (02)	.65 (03)		.12 (03)	.45 (02)
3				.76 (03)			.42 (03)
4						.85 (03)	.28 (03)
5	-.19 (04)		-.08 (03)			.85 (04)	.40 (03)
6		.12 (03)	.08 (03)			.44 (03)	.75 (03)
7	.17 (03)	.59 (03)	-.10 (03)			.10 (03)	.55 (03)
8		.80 (03)					.36 (02)
9		.93 (03)					.14 (02)
10			.65 (03)	-.08 (03)	.18 (03)		.51 (03)
11			.95 (03)				.10 (03)
12			.63 (03)				.60 (03)
13					.82 (03)		.32 (03)
14	.16 (03)		-.22 (03)	.11 (03)	.62 (04)		.52 (03)
15	.13 (03)			.11 (03)	.70 (03)		.41 (02)
16	.82 (03)		.14 (03)		.28 (03)		.26 (02)
17	.72 (03)	.18 (03)	.12 (03)		.22 (03)		.35 (02)
18	.74 (03)						.45 (03)

Note: N = 1244; standard errors of parameters estimates are within parentheses, decimals have been left off; nonsignificant factor loadings are left blank and fixed at zero for all subsequent analyses.

Table 4

Estimated Factor Variance-Covariance Matrix for the Eighteen Subscales

	Factor					
Factor	I	II	III	IV	V	VI
I	1.00					
II	.07 (04)	1.00				
III	-.35 (04)	.19 (04)	1.00			
IV	-.05 (04)	.21 (03)	.20 (03)	1.00		
V	.06 (05)	.47 (03)	.28 (04)	.45 (03)	1.00	
VI	.49 (04)	.24 (03)	-.03 (04)	.29 (03)	.36 (03)	1.00

Note: Standard errors of parameter estimates are within parentheses, decimal points are omitted.

Table 5

Maximum Likelihood Factor Analyses and Simultaneous Factor Analyses for American and Japanese Cultures

MAXIMUM LIKELIHOOD FACTOR ANALYSIS (Unconstrained Model)

Group	N	χ^2_{98}	GFI	AGFI	RMR
American	746	244.11	.950	.913	.027
Japanese	498	162.99	.945	.904	.028
		407.10			

SIMULTANEOUS FACTOR ANALYSIS (Constrained Models)

Goodness of Fit Indices	Invariant Matrix		
	Λ_y & Φ^2	Λ_y	Φ^2
χ^2, df	570.16, 242	501.09, 229	470.34, 212
$\Delta\chi^2$, Δdf	163.06, 46	93.99, 33	63.24, 16
Ratio	3.54	2.84	3.96
GFI			
American	.929	.935	.945
Japanese	.902	.914	.934
RMR			
American	.036	.034	.029
Japanese	.043	.038	.032

Table 6

Estimated Factor Loadings of Subscales Resulting From Separate Maximum Likelihood Factor Analyses of Rescaled Variance-Covariance Matrices for American and Japanese Cultures

Group	Subscale	Factor I	II	III	IV	V	VI	D^2
American	1				.84 (03)			.29 (03)
	2		.04 (03)	-.00 (03)	.68 (04)		.12 (04)	.46 (03)
	3				.75 (04)			.43 (03)
	4						.86 (04)	.27 (03)
	5	-.13 (05)		-.04 (03)			.83 (05)	.41 (04)
	6		-.00 (04)	.03 (03)			.55 (04)	.70 (04)
	7	.05 (04)	.67 (04)	-.06 (03)			.10 (04)	.48 (03)
	8		.79 (03)					.38 (03)
	9		.91 (03)					.17 (03)
	10			.60 (04)	-.05 (04)	.14 (04)		.60 (04)
	11			.96 (04)				.07 (05)
	12			.59 (04)				.65 (04)
	13					.77 (04)		.40 (04)
	14	.09 (04)		-.19 (03)	.05 (04)	.68 (05)		.46 (03)
	15	.07 (04)			.10 (04)	.72 (04)		.37 (03)
	16	.81 (04)		.11 (04)		.16 (04)		.26 (03)
	17	.72 (04)	.14 (03)	-.00 (04)		.16 (04)		.30 (02)
	18	.73 (03)						.47 (03)
Japanese	1				.82 (04)			.33 (04)
	2		.07 (04)	.22 (04)	.61 (04)		.14 (04)	.42 (03)
	3				.76 (04)			.42 (04)
	4						.77 (04)	.42 (04)
	5	-.40 (07)		-.09 (05)			.98 (08)	.24 (06)
	6		.04 (05)	-.02 (05)			.49 (05)	.75 (05)
	7	.12 (05)	.65 (05)	-.04 (04)			.05 (05)	.49 (04)
	8		.77 (04)					.41 (04)
	9		.91 (04)					.18 (04)
	10			.65 (04)	-.07 (05)	.10 (05)		.55 (04)
	11			.95 (04)				.09 (04)
	12			.71 (04)				.50 (04)
	13					.81 (04)		.34 (04)
	14	-.03 (05)		-.16 (04)	.03 (07)	.78 (07)		.42 (04)
	15	.11 (05)			.05 (06)	.68 (06)		.46 (04)
	16	.73 (04)		.02 (05)		.27 (05)		.31 (03)
	17	.78 (05)	.00 (05)	.12 (05)		.16 (05)		.30 (03)
	18	.72 (04)						.49 (04)

Note: Standard errors of parameter estimates are within parentheses, decimal points are omitted.

Table 7

Estimated Factor Loadings of Items Resulting From Separate Maximum Likelihood Factor Analyses of Rescaled Variance-Covariance Matrices for American and Japanese Cultures

Group	Item	Factor I	II	III	IV	V	VI	h^2
American	1		.10 (05)	.05 (04)	.28 (04)		.20 (04)	.80 (04)
	2		-.07 (03)	-.06 (03)	.71 (05)		.11 (04)	.48 (04)
	3		-.04 (05)	-.01 (04)	.53 (05)		-.02 (04)	.74 (04)
	4	-.20 (06)		-.09 (04)			.81 (05)	.51 (03)
	5	-.30 (06)		-.03 (04)			.93 (05)	.39 (04)
	6			.26 (04)				.93 (05)
	7			.59 (04)				.64 (04)
	8	.62 (05)	.10 (05)	-.02 (04)		.00 (05)		.58 (04)
	9	.49 (04)	.15 (05)	.03 (04)		.22 (05)		.59 (04)
	10	.32 (05)	.14 (05)	-.09 (04)		.24 (05)		.73 (04)
Japanese	1		.28 (06)	.23 (05)	.25 (05)		.04 (05)	.73 (05)
	2		-.19 (06)	.00 (04)	.68 (06)		.18 (05)	.45 (05)
	3		.06 (06)	.19 (05)	.29 (05)		.13 (05)	.79 (05)
	4	-.42 (07)		-.08 (05)			.98 (06)	.35 (04)
	5	-.33 (06)		-.06 (05)			.90 (06)	.46 (04)
	6			.35 (05)				.85 (06)
	7			.72 (05)				.48 (04)
	8	.72 (06)	-.06 (09)	.12 (05)		.00 (07)		.53 (04)
	9	.70 (05)	.04 (08)	.10 (05)		.20 (07)		.40 (04)
	10	.55 (05)	-.08 (09)	.02 (05)		.34 (08)		.58 (05)

Note: Standard errors of parameter estimates are within parentheses, decimal points are omitted.

Table 8

Significance Levels for Pairwise Comparison of Item Factor Loadings Across American and Japanese Cultures

Subscale	Item	$\rho(X_i, L_i)$	Significance Level
	1	.935	.000 **
2	2	.874	.443
	3	.893	.000 **
5	4	.953	.051 *
	5	.932	.757
12	6	.938	.131
	7	.957	.032 *
	8	.953	.156
17	9	.957	.028 *
	10	.954	.006 **

Footnotes

[1] These samples are a part of the data base established for the study of Personal Investment. The American sample is described in Braskamp and Maehr (1983). The Japanese sample was collected by David Shwalb of the University of Michigan. Appreciation for the use of these data is acknowledged.

REFERENCES

Braskamp, L. A. and Maehr, M. L. Development of the Inventory of Personal Investment. Paper presented at the annual meeting of the American Educational Research Association, Montreal, April, 1983.

Box, G. E. A general distribution theory for a class of likelihood criteria. Biometrika, 1949, 36, 317-346.

Christopher, R. C. The Japanese Mind: The Goliath Explained. New York: Linden Press, 1983.

Drasgow, F. and Dorans, N. J. Robustness of estimators of the squared multiple correlation and squared cross-validity coefficient to violations of multivariate normality. Applied Psychological Measurement, 1982, 6, 185-200.

Drasgow, F. and Kanfer, R. Equivalence of psychological measurement in hetergeneous populations. Manuscript in preparation, 1983.

Hackman, J. R. and Oldham, G. R. Motivation through the design of work: Test of a theory. Organizational Behavior and Human Performance, 1976, 16, 250-279.

Hayashi, C. Changes in belief systems, quality of life issues, and social conditions over 25 years of post-war Japan. Paper presented at the 10th World Congress of Sociology, Mexico City, August, 1982.

Hulin, C. L., Drasgow, F., and Komocar J. Applications of item response theory to analysis of attitude scale translations. Journal of Applied Psychology, 1982, 67, 818-825.

Ironson, G. H. and Subkoviak, M. J. A comparison of several methods of assessing item bias. Journal of Educational Measurement, 1979, 16, 209-225.

Irvine, S. H. and Carroll, W. K. Testing and assessment across cultures: Issues in methodology and theory. In H. C. Triandis and J. W. Berry (Eds.) Handbook of Cross-Cultural Psychology (Vol 2.) Boston: Allyn and Bacon, 1980.

Joreskog, K. G. Simultaneous factor analysis in several populations. Psychometrika, 1971, 36, 409-426.

Joreskog, K. G. Some contributions to maximum likelihood factor analysis. Psychometrika, 1967, 32, 443-482.

Joreskog, K. G. and Sorbom, D. Advances in Factor Analysis and Structural Equation Models. Cambridge, MA: Abt Books, 1979.

Joreskog, K. G. and Sorbom, D. LISREL V: Analysis of Linear Structural Relationsips by the Method of Maximum Likelihood User's Guide. Chicago: National Educational Resources, 1981.

Katerberg, R., Hoy, S., and Smith, F. J. Language, time, and person effects on attitude scale translation. Journal of Applied Psychology. 1977, 62, 385-391.

Kroonenberg, P. M. and Lewis, C. Methodological issues in the search for a factor model: Exploration through confirmation. Journal of Educational Statistics, 1982, 7, 69-89.

Lawley, D. N. A note on Karl Pearson's selection formulae. Proceedings of the Royal Society of Edinburgh, Section A, 1943-44, 62, 28-30.

Lord, F. M. Applications of Item Response Theory to Practicial Testing Problems. Hillsdale, NJ: Erlbaum, 1980.

Maehr, M. L. Meaning and motivation: Toward a theory of Personal Investment. In R. Ames and C. Ames (Eds.) Research on Motivation in Education: Student Motivation. New York: Academic Press, 1984.

McGaw, B. and Joreskog, K. G. Factorial invariance of ability measures in groups differing in intelligence and socioeconomic status. British Journal of Mathematical and Statistical Psychology, 1971, 24, 154-168.

Meredith, W. Rotation to achieve factorial invariance. Psychometrika, 1964, 29, 187-206.

Mulaik, S. A. The Foundations of Factor Analysis. New York: McGraw-Hill Book Company, 1972.

Ouchi, W. G. Theory Z: How American Business Can Meet the Japanese Challenge. New York: Avon Books, 1981.

Scheumeman, J. A method of assessing bias in test items. Journal of Educational Measurement, 1979, 16, 143-152.

Tatsuoka, M. M. Multivariate Analysis: Techniques for Educational and Psychological Research. New York: John Wiley and Sons, 1971.

Thurstone, L. L. Multiple Factor Analysis. Chicago: University of Chicago Press, 1947.

Wood, R. L., Wingersky, M. S., and Lord F. M. LOGIST: A computer program for estimating examinee ability and item characteristic curve parameters. Research Memorandum 76-6. Princeton, NJ: Educational Testing Service, 1976.

Wright, B. D. Solving measurement problems with the Rasch model. Journal of Educational Measurement, 1977, 14, 91-116.